PATENT APPLICATION

SPHINGOSINE 1-PHOSPHATE RECEPTOR GENE, SPPR

Inventors: Thomas P. Loughran, Jr. & Ravi Kothapalli

FIELD OF THE INVENTION

The present invention relates to the genetics of autoimmune diseases, including lymphoproliferative diseases, such as large granular lymphocyte leukemia (LGL), and rheumatoid arthritis (RA). Specifically, the invention relates to a novel sphingosine 1-phosphate receptor gene, herein termed sppr, and its splice variants. Sppr is up-regulated in LGL and is useful, for example, in the diagnosis and treatment of certain lymphoproliferative, neurodegenerative and autoimmune diseases.

BACKGROUND OF THE INVENTION

Large granular lymphocyte leukemia (LGL) is a rare form of lymphoproliferative disorder often associated with autoimmune disease (Loughran T. P., Clonal diseases of large granular lymphocytes. Blood 82, 1-14, 1993).

The cause of LGL is still not fully understood. An increased count of large granular lymphocytes is characteristic of LGL leukemia. Patients with clonal CD3+LGL, as determined by cytogenetic or T-cell receptor (TCR) gene rearrangement studies, are classified as T-LGL. Some of these patients may resemble those with Felty's syndrome with clinical features of rheumatoid arthritis, neutropenia and splenomegaly (Ahern M. J., et al., P. Phenotypic and genotypic analysis of mononuclear cells from patients with Felty's syndrome. Ann.Rheum.49, 103-108, 1990.) Morbidity and mortality in patients with LGL leukemia typically results from infections acquired during severe neutropenia.

The etiology of LGL leukemia is also not yet known. There is strong evidence that suggests that leukemic large granular lymphocytes are antigen activated cytotoxic T lymphocytes (CTL), but the nature of the antigen and of the initial stimulus leading to antigen driven expansion are not known.

LGL leukemic cells express FAS and FAS ligand, but they are not actively undergoing apoptosis (Perzova, R and Loughran, T. P, Jr. Constitutive expression of Fas

ligand in large granular lymphocyte leukemia. British Jnl. Haematology, 1997). How they acquire resistance to apoptosis is not known.

Within the field of the diagnosis and treatment of LGL and other autoimmune diseases, there is a need for better tools for diagnosis and early detection of disease, specific therapeutic targets and treatments for the disease, and more specific reagents and tools with which to identify the pathogenic pathways of these diseases. The present invention provides a novel gene and splice variants that are linked to these diseases, and which address the aforementioned needs and more, as will become clear to one of skill in the art upon reading the following disclosure.

SUMMARY OF THE INVENTION

Large granular lymphocyte leukemia (LGL) is a lymphoproliferative disorder often associated with autoimmune disease. In order to identify differentially expressed genes in LGL leukemia, microarray analysis is performed from RNA isolated from PBMC of LGL leukemia patients and compared with normal healthy individual(s). By screening a human LGL leukemia library the full-length sequence of a human gene that showed 85% identity with rat sphingosine l-phosphate receptor is obtained. Two different isoforms are also identified by RT-PCR, designated sphingosine 1-phosphate receptor 1, also referred to as S1p5-α and sphingosine 1-phosphate receptor 2, also refered to as S1P5-β. Sphingosine 1-phosphate receptor *(sppr)* is present in brain, spleen, PBMCs, liver and kidney. The present inventors found it is over-expressed in LGL leukemia patients when compare to normal individuals.

In a first embodiment, the invention provides a gene comprising sppr or a splice 5 variant, or sppr protein or modified proteins or fragments thereof.

In a further embodiment, the invention provides a nucleic acid capable of hybridizing to at least a portion of said sppr gene, including splice variants.

In a further embodiment, the invention provides methods for screening for autoimmune diseases, including LGL or rheumatoid arthritis, based on overexpression of sppr.

In a further embodiment, the invention provides for monoclonal antibodies to sppr and their use in detection, diagnosis and treatment of disease states.

In a further embodiment, the invention provides for screening of ligands, agonists, and antagonists of sppr.

In a further embodiment, the invention provides for inhibition or treatment of neurodegenerative disease.

In a preferred embodiment the present invention provides a sphingosine 1-phosphate receptor gene. The use of said gene makes it possible to produce the sphingosine 1 -phosphate receptor protein with ease and in large quantities, and said protein, which has sphingosine 1-phosphate receptor activity, can be used in developing therapeutic agents for various diseases.

Throughout this document the nomenclature sppr and S1P5 are used interchangeably. The receptor was initially termed sppr. However, to be consistent with a new nomenclature system this receptor was renamed S1P5.

DESCRIPTION OF THE FIGURES

FIGURE lA-B illustrates a microarray of the differential expression of the selected EST. (EST (GenBank ID 1868427) is obtained Incyte Genomics.) Figure 1A-B shows a microarray hybridized with the fluorescent labeled probes generated using mRNA isolated from PBMC of LGL leukemia patient and from mRNA isolated from normal healthy individual. Figure 1A illustrates a microarray showing the expression of an LGL leukemia patient cDNAs. Figure 1B illustrates a microarray showing the expression of a normal healthy individual. Arrows show the expression of EST in both patient and normal individual (GeneBank Id: N47089). Intensity bar shows the increased expression starting from left to right. A balanced differential expression of 3.0 is determined for this EST.

FIGURE 2 shows Northern blot analysis performed with 10 ug of total RNA isolated from PBMC of LGL leukemia patients and normal healthy individuals. These results demonstrate over-expression of EST in the PBMCs of LGL leukemia when compared to normal and normal activated PBMCs of healthy individuals.

FIGURE 3 shows the complete nucleotide sequence, SEQ ID NO: 4, of human sphingosine 1-Phosphate receptor (SPPR) cDNA and amino acid sequence (SEQ ID NO: 3) as predicted by the nucleic acid sequence. The full-length (2.2 kb) nucleotide sequence of *sppr* is compiled from sequences of clones isolated from an LGL leukemia library and RT-PCR products obtained by using the gene specific primers designed using the corresponding sequence from chromosome 19.

FIGURE 4 shows the alignment of the amino acid sequence of SPPR with other members of the sphingosine 1-phosphate receptor family. The deduced amino acid sequence of *sppr* is compared with rat edg-1 and nrg-1 There is approximately 85% identity with these genes.

FIGURE 5 shows the nucleotide sequence and deduced amino acid sequence of splice variant, sphingosine 1-phosphate receptor l. 1.6 kb fragment is obtained by RT-PCR using total RNA isolated from PBMC of an LGL leukemia patient. The fragment is then cloned and sequenced.

FIGURE 6 shows the nucleotide sequence and deduced amino acid sequence of splice variant, sphingosine 1-phosphate receptor 2. The nucleotide sequence of an alternative splice variant of *sppr* and deduced sequence. 1.2 kb fragment is obtained from RT-PCR using total RNA isolated from PBMC of LGL leukemia. The fragment is then cloned and sequenced.

FIGURE 7 shows results of *sppr* Northern blot analysis with different tissues. Northern blot analysis is performed using a multiple tissue Northern blot (Clontech). Northern blots contain approximately l ug of poly A+ per lane from twelve different human tissues. A 1 .5 kb fragment containing the full-length open reading frame for *sppr* is used as a probe. Results show *sppr* is expressed in mainly brain, spleen, and peripheral blood leukocytes. Small amounts of *sppr* are also expressed in lung, placenta, liver and kidney.

FIGURE 8 shows nucleotide and deduced amino acid sequence of human $S1P_5$ cDNA. Full-length (2.2 kb) nucleotide sequence of $S1P_5$ is compiled from the sequences of clones isolated from LGL leukemia library (clone 6) and RT-PCR products. GenBank Accession No. AF331840. The predicted amino acids of the coding region are shown underneath by a single letter abbreviation. The left side of the sequence shows nucleotide

numbers and the right side shows amino acid numbers. Possible seven transmembrane helices are underlined. The putative polyadenylation sites are in bold.

FIGURE 9 shows Alignment of the deduced amino acid sequence of *S1P5* with other members: The deduced amino acid sequence of *S1P5* is compared with predicted amino acid sequences of rat edg-8 and nrg-1. There is approximately 86% identity with these genes. * - single, fully conserved residue, : - conservation of strong groups, . - conservation of weak groups, - no consensus.

FIGURE 10 shows activation of Erk2 by S1P in HEK293 cells transiently transfected with *S1P5*. HEK 293 cells transfected with the HA-ERK2 plasmid (0.2 μg) and either pcDNA *S1P5* (0.5 μg) or vector alone. Vector plasmid is added to each transfection reaction to equalize the amount of total DNA (2.1 μg). After serum-starvation, the cells are treated with 1μM S1P or 1μM LPA for 5 min (BSA was added to the controls). HA-ERK2 is immunoprecipitated from one half of each whole cell lysate and used for measuring the kinase activity utilizing MBP as substrate, while HA-ERK2 immunoprecipitated from the other half is used for determining the amount of ERK2 protein in the immune complex. Figure 10 illustrates a representative autoradiogram of ^{32}P incorporation into MBP catalyzed by HA-ERK 2 immunoprecipitated from transiently transfected cells treated as indicated. Figure 10 further illustrates the corresponding Western blot demonstrating the amount of HA-ERK2 present in each of the immune complexes. Figure 10 further illustrates a plot of ERK 2 activity (fold) normalized to the amount of ERK2 protein (means ± SD from three independent experiments).

FIGURE 11 shows Northern blot analysis of *S1P5* mRNA expression in PBMC of LGL leukemia patients and normal healthy individuals. Northern blot is performed with 10 μg of total RNA isolated from PBMC of LGL leukemia patients and normal healthy individuals. LGL = LGL leukemia patients, N= Normal healthy individual NA= Normal healthy individuals PBMCs activated by IL2 and PHA. These results demonstrate over-expression of *S1P5* in the PBMC of LGL leukemia when compared to normal and normal activated PBMC of healthy individuals.

FIGURE 12A-B shows comparison of the predicted amino acid sequences of *S1P5*, *S1P5*-α and *S1P5*-β. The predicted amino acid sequences are aligned using CLUSTAL

program. Figure 12A illustrates the nucleotide sequence of an alternative splice variant of *S1P*$_5$-α and deduced amino acid sequence. A 1.6 kb fragment is obtained from RT-PCR using total RNA isolated from PBMC of LGL leukemia patient. This fragment is cloned and sequenced. Figure 12B illustrates the nucleotide sequence of an alternative splice variant of *S1P*$_5$-β and deduced sequence. A 1.2 kb fragment is obtained from RT-PCR using total RNA isolated from PBMC of LGL leukemia. This fragment is cloned and sequenced.

FIGURE 13 shows tissue distribution of *S1P*$_5$ message. Northern blot analysis is performed using the multiple tissue blot obtained from Clontech. The Northern Blot contains approximately 2 μg of poly A+ per lane from twelve different human tissues and a 1.5 kb fragment containing the full-length open reading frame of *S1P*$_5$ is used as a probe. As shown above, *S1P*$_5$ is expressed mainly in brain, spleen, and peripheral blood leukocytes. Trace amounts of *S1P*$_5$ are also expressed in lung, placenta, liver and kidney. (Please note: Signals are significantly stronger in normal tissue on poly A + RNA Northern blot compared to total RNA Northern blot.)

DETAILED DESCRIPTION OF THE INVENTION

The abbreviations for amino acids, peptides, base sequences, nucleic acids and so forth as used herein in the present specification are those recommended by the International Union of Pure and Applied Chemistry (IUPAC) and the International Union of Biochemistry (IUB) and in the "Guidelines for drafting patent specifications relative to base sequences and/or amino acid sequences" edited by the Japanese Patent Office or those commonly used in the relevant field of art.

Although the genes of the present invention is represented by a single-stranded DNA sequence, as shown under, for example, SEQ ID NO:4, the present invention also includes the DNA sequence complementary to such a single-stranded DNA sequence as well as a component comprising both of these. The DNA sequence representing the gene of the present invention shown in the above-mentioned SEQ ID NO: 4 is an example of the codon combination coding for the respective amino acid residues according to the amino acid sequence shown in SEQ ID NO:7. The gene of the present invention is not limited to the above-mentioned one but may, of course, have any other DNA base sequence comprising a combination of codons arbitrarily selected for the respective amino acid residues without

altering the above-mentioned amino acid sequence. Selection of said codons can be carried out by the conventional method in which the codon usage or codon choice in the host to be used for gene recombination is taken into consideration [Nucl. Acids Res., 9,43-74(1981)], and these codons can be produced, for example by chemical synthesis, etc.

The gene of the present invention further includes DNA sequences coding for those equivalents to the above-mentioned amino acid sequence that are derived from the latter by deletion, addition or like modification of one or more amino acid residues or part of the amino acid sequence and have similar sphingosine l-phosphate receptor activity to that of the sphingosine l-phosphate receptor protein. While production, alteration (mutation) or the like of these polypeptides may occur spontaneously, they can also be produced by posttranslational modification. Furthermore, any desired gene can be produced by gene engineering techniques such as the site-specific mutagenesis technique in which the natural gene (gene of the present invention) is altered, by a chemical synthesis technique such as the phosphite triester method in which mutant DNAs are synthesized or by combining both procedures. By utilizing the gene of the present invention, namely by incorporating the same into a vector for use with a microorganism, for instance, and cultivating the transformant microorganism, the sphingosine 1-phosphate receptor protein can be expressed readily and in large quantities, and said protein can be isolated and provided. Since said protein has sphingosine 1-phosphate receptor activity, it is effective for various pharmacological purposes, and it is also useful, among others, in elucidating the pathogenesis, the pathologies or the like of various diseases. More specifically, the recombinant sphingosine 1-phosphate receptor protein obtained by utilizing the gene of the present invention can effectively be used, for example, in elucidating the mechanism of immunosuppression in living bodies, developing or screening out therapeutic agents for autoimmune diseases (e.g. rheumatism, SLE (systemic lupus erythematodes), LGL, etc.), searching for endogenous ligands and substrates to the novel protein and developing therapeutic agents therefor.

Similarly, the gene of the present invention can effectively be used, for example, in elucidating the mechanism of neurodegeneration in living bodies, developing or screening out therapeutic agents for neurodegenerative diseases (e.g. alzheimers, parkinson's and the like), searching for endogenous ligands and substrates to the novel protein and developing therapeutic agents therefor.

In the following, the gene of the present invention will be described in more detail. The gene of the present invention can be isolated by general genetic engineering techniques, for example, by selecting an appropriate clone from among a human fetal brain cDNA library (cDNA synthesized in the conventional manner from mRNA isolated and purified from total RNA obtained in turn from appropriate origin cells containing a gene coding for the sphingosine 1-phosphate receptor protein) using appropriate probes, such as for example those of SEQ ID 1 and SEQ ID2, purifying said clone, and determining the base sequence thereof. In the above procedure, the origin cells may be any animal cells or tissues where the occurrence of sphingosine 1-phosphate receptor protein is known (see for example, the experiment producing the results shown in FIG 6), or soluble fractions of cultured cells derived therefrom. This can be isolated and purified for the culture supernatant by various chromatographic processes.

In the practice of the present invention, it is also possible to use a part of the DNA fragment sequenced in the above manner as a probe, label this using a random prime DNA labeling kit (available from Takara Shuzo, Amersham, etc.) in accordance with the random prime DNA labeling method (Feinberg, A. P., et al., Anal. Biochem., 137, 266-267 (1984)), for instance, and use the thus-obtained labeled probe in screening out the desired sphingosine l-phosphate receptor protein gene.

Using the above-mentioned labeled probe, for instance, the desired DNA can be screened out by the plaque hybridization technique developed by Benton and Davis (Benton, W. and Davis, R., Science, 196, 383-394 (1977)).

The gene of the present invention as obtained in the above manner can be cloned in various plasmids in the conventional manner. For instance, after cleavage with an appropriate restriction enzyme and purification, the gene of the present invention can be inserted into a cloning vector (e.g. plasmid) cleaved with the same restriction enzyme and purified, at the cleavage site thereof, whereby a recombinant plasmid can be obtained. By introducing said recombinant into an appropriate host (e.g. Escherichia coli) for transformation, a restriction enzyme map of the clone containing said gene can be drawn using the transformant by a conventional known method, for example the method as described in Sambrook, J. Fritsch, E.F., and Maniatis. Molecular cloning. A laboratory Manual 2nd edition. Cold Spring Harbor laboratory Press. Cold Spring Harbor, NY. After digestion of the above clone with an

appropriate restriction enzyme, the base sequence of said clone can be determined by the above-mentioned dideoxy method or the Maxam-Gilbert method, for instance. The base sequence determination mentioned above may also be readily performed using a commercially available kit or the like.

The thus-determined DNA base sequence of the sphingosine 1-phosphate receptor protein gene of the present invention and the corresponding amino acid sequence encoded thereby are as shown in the sequence listing under SEQ ID NO: 3 and SEQ ID NO:4.

Using the above-mentioned gene (DNA) of the present invention, the recombinant sphingosine 1-phosphate receptor protein can be obtained by various known gene recombination techniques [cf. for example Science, 224, 1431 (1984); Biochem. Biophys. Res. Comm., 130, 692 (1985); Proc. Natl. Acad. Sci. USA, 80, 5990 (1983)]. Said sphingosine 1-phosphate receptor protein is produced, in more detail, by constructing a recombinant DNA allowing expression of the gene of the present invention in host cells, introducing this into host cells for transformation thereof, and cultivating the transformant strain. The host cells may be either eukaryotic or prokaryotic. As an expression vector for use with vertebrate cells, it is possible to use one containing a promoter generally located upstream of the gene to be expressed, an RNA splicing site, a polyadenylation site and a transcription termination sequence and so on. This may further have a replication origin, as necessary. Yeasts are often and generally used as eukaryotic microorganisms and, among them, yeasts belonging to the genus Saccharomyces are advantageously used. Usable as expression vectors for use with said yeasts and other eukaryotic microorganisms are pAM82 (A. Miyanohara et al., Proc. Natl. Acad. Sci. USA, 80, 1-5 (1983)) containing a promoter for the acid phosphatase gene, and like vectors. Escherichia coli and Bacilus subtilis are generally and very often used as prokaryotic host cells. When these are used as hosts in the practice of the present invention, an expression plasmid is preferably used which is derived, for instance, from a plasmid vector capable of replication in said host microorganisms and provided with a promoter, the SD (Shine and Dalgarno) base sequence and further an initiation codon (e.g. ATG) necessary for the initiation of protein synthesis, upstream from the gene of the present invention so that said gene can be expressed. As the host Escherichia coli mentioned above, the strain Escherichia coli K12 and the like are often used and, as the vector, pBR322 is generally and often used. However, the host and vector are not limited

thereto, but other various known microbial strains and vectors can also be used. As regards the promoter, the tryptophan (trp) promoter, l pp promoter, lac promoter and P.sub.L promoter, for instance, can be used.

The thus-obtained desired recombinant DNA can be introduced into host cells for transformation thereof by various conventional methods. The transformant obtained can be cultivated in the conventional manner, leading to production and accumulation of the desired sphingosine l-phosphate receptor protein encoded by the gene of the present invention. The medium to be used in said cultivation can adequately be selected, according to the host cells employed, from among various media in common use. When Escherichia coli or like cells are used as host cells, for instance, transformant cultivation can be conducted using LB medium, E medium, M9 medium, M63 medium or the like. To these media, there may be added, as necessary, generally known various carbon sources, nitrogen sources, inorganic salts, vitamins, nature-derived extracts, physiologically active substances, etc. The above-mentioned transformant cultivation can be carried out under conditions suited for the growth of the host cells. In the case of Escherichia coli, such conditions can be employed, for instance, as a pH of about 5 to 8, preferably 7 or thereabout, and a temperature of about 20 to 43.degree. C., preferably 37.degree. C. or thereabout. In the above manner, the transformant cells produce and accumulate intracellularly or secrete extracellularly the desired recombinant FK506 binding protein.

Said desired protein can be isolated and purified by various separation techniques utilizing its physical, chemical and other properties [cf. for example "Seikagaku (Biochemistry) Data Book II", pages 1175-1259, 1st edition, 1st printing, published Jun. 23, 1980 by Kabushiki Kaisha Tokyo Kagaku Dojin; Biochemistry, vol. 25, No. 25, 8274-8277 (1986); Eur. J. Biochem., 163, 313-321 (1987)]. As specific examples of said techniques, there may be mentioned conventional reconstitution treatment, treatment with a protein precipitating agent (salting out), centrifugation, osmotic pressure shock treatment, ultrasonic disruption, ultrafiltration, various liquid chromatographic processes such as molecular sieve chromatography (gel filtration), adsorption chromatography, ion exchange chromatography, affinity chromatography and high performance liquid chromatography (HPLC), dialysis, and combinations of these. In the above manner, the desired recombinant protein can be produced on an industrial scale with ease and with high efficiency.

In order to provide diagnostics for LGL leukemia, and provide therapeutic targets for drugs directed to mitigate the pathogenesis of LGL leukemia, microarray analysis is performed to identify differentially expressed genes. A large number of genes are identified that are differentially expressed in LGL leukemia compared to normal controls. One of the ESTs of approximately 300 base pairs is fully characterized herein. Initial Blast analysis shows 100 % homology with Homo-sapiens full-length insert cDNA clone YY 85D04 (gb/AF 088014). No open reading frame within the full-length insert cDNA. Therefore, in order get the complete sequence of the gene, the LGL leukemia library is screened and also RT-PCR is performed using the total RNA isolated from different LGL leukemia patients. 15 positive clones are selected from library screening. All of them give partial sequences with the longest one being approximately 340 base pairs shorter (clone 6). BLAST search with htgs, shows that clone 6 shows 100% homology with genomic sequence present in the chromosome 19. Primers are designed based on the genomic sequence information to obtain full-length sequence of the gene. By using these primers in the PCR with genomic DNA and RT-PCR with total RNA, the full-length gene, SEQ ID:4 is obtained. This gene belongs to the G-protein-coupled receptor super-family of integral membrane proteins. BLAST analysis of the complete gene reveals 85% homology with rat sphingosine 1-phosphate receptor *edg-8 and nrg-1(*Im, D., et al., Characterization of a Novel sphingosine 1-Phosphate receptor, Edg-8. J.Biol.Chem.275. 1428 1-14286 (2000); Glickman, M., et al., Molecular cloning, tissue-specific expression and chromosomal localization of a novel nerve growth factor regulated G-protein-coupled receptor, nrg-1 . Mol. Cell. Neurosci. 14, 141-152 (1999)), shown in FIG.4. It is interesting to note that this gene is present mainly in brain, spleen and PBMCs (FIG.7), and it is over expressed in PBMC of LGL leukemia patients and is be involved in LGL leukemia cell survival or proliferation.

Material and Methods:

Isolation of Peripheral blood mononuclear cells *(PBMC and RNA).* PBMC are isolated from normal healthy individuals and from LGL leukemia patients. Trizole is obtained from GTBCO-BRL. EST (GenBank ID 1868427) is obtained Incyte Genomics. Oligotex mRNA mini-kit, plasmid isolation kits, gel extraction kits, and PCR reagents are purchased from Qiagen; RNA loading dye is from Sigma Chemical Co. The Prime-a-Gene labeling kit is from Promega Corp. (Madison, WI). Deoxycytidine 5'triphosphate dCTP a-

32P (3,000 Ci/ mmol) is from Dupont NEN (Boston, MA). Nytran membrane is obtained from Schleicher& Schuell, Inc., 10 optical Avenue, Keene, N.H. Nick translation columns are obtained from Pharmacia Chemical Co. The Topo-TA cloning kit is from Invitrigen.

PBMC are isolated from whole blood using Ficoll-Hypaque density gradient centrifugation. The PBMC cells are suspended in Trizole reagent (GIBCO-BRL, Rockville, MD) and total RNA is immediately isolated according to the Oligotex mRNA mini-kit manufacturer's instructions and stored at —70 ^{0}C. Poly A+ RNA is isolated from total RNA by using Oliogo-Tex mini mRNA kit according to the manufacturer's recommendations. PBMCs are cultured *in vitro* and activated by Interleukin 2 and phytohemagglutinin (PHA) for 2 to 3 days. In a preferred embodiment, PBMC is cultured *in vitro* and activated by PHA, (Sigma Chemical Co. St. Louis, MO) (1μg/ml, 2 days) and Interleukin-2 (IL-2) (100 U/ml, 10 days), Next, total RNA is isolated as described above.

Microarray probing and analysis is done by Incyte Genomics, (St. Louis, Missouri). Approximately 1 ug of Poly (A)$^{+}$ RNA isolated from PBMCs of LGL leukemia and healthy individual is reverse transcribed to generate Cys3 and Cys 5 fluorescently labeled cDNA probes. In a preferred embodiment, more than 90% of PBMC from the LGL leukemia patient are leukemic LGL as indicated by CD 8^{+} staining. cDNA probes are competitively hybridized to a human UniGEM V cDNA microarray containing approximately 7075 immobilized cDNA fragments (4107 known genes and 2968 ESTs). Scanning and quantitation is performed by Incyte Genomics and balanced differential differentiation is given for all the genes. The balanced differential expression is calculated using the ratio between the P1 signal (intensity reading for probe 1) and the balanced P2 signal (intensity reading for probe 2 adjusted using the balanced coefficient). A balanced differential expression of 2.0 is considered indicative of up-regulation of a given gene.

Verification of clones: GEM cDNA clones are purchased from Incyte Genomics as individual bacterial stabs and streaked on LB /agar plates containing appropriate antibiotic(s). Individual colonies are picked and grown in LB medium. Plasmid DNA is isolated and sequenced in order to verify the correct identity of each clone.

Northern Blot analysis: Northern Blotting is done as described previously (Sambrook et al, 1998). Essentially, 10 ug of total RNA from each sample is denatured at 65 ^{0}C in a

RNA loading buffer, electrophoresed in 1% agarose containing 2.2 M formaldehyde gel, and blotted onto a Nytran membrane. (Nytran membrane obtained from Schleicher & Schuell, Inc, Keene,N.H). The RNA is fixed to the membrane by UV cross-linking. cDNA is labeled with [^{32}P] (Prime-a-Gene labeling kit from Promega Corp. Madison, WI, deoxycytidine 5'triphosphate (dCTP α-^{32}P, 3,000 Ci/ mmol, Dupont NEN, Boston, MA) and purified by Nick columns (Amersham Pharmacia Biotech AB, Piscataway, NJ). Hybridization and washings of the blots are performed as described by Engler-Blum, G., Meier, M ., Frank, J., and Muller, G.A. Reduction of background in problems in non-radioactive Northern blot analysis enables higher sensitivity than 32P-based hybridizations. Anal. Biochem. 210, 235-244 (1993).

Library construction and screening. cDNA is synthesized from poly(A)$^+$ RNA isolated from pooled PBMCs of multiple LGL leukemia patients using oligo dT primer. The cDNA is unidirectionally inserted the EcoRI / XhoI sites of Lambda ZAPII (Stratagene).cDNA library is screened using EST according to standard protocol (Sambrook *et al.*,1989). In a preferred embodiment, DNA libraries are plated at a density of 50,000 plaque-forming units per 150mm plate. Following incubation for 8 h at 37 ^{0}C, the plated phage are overlaid with nitrocellulose filters. After 1 min the filters are removed and the membranes are crossed linked by Autocross linker. A [^{32}P] labeled cDNA fragment derived from an EST (GenBank accession No. N 47089) of interest is used to probe the filters. Hybridizations, washings, exposure of the membranes to films and then picking up the colony of interest are performed as outlined in the standard methodology (Sambrook *et al.,* 1989). Secondary and tertiary screenings were also performed as outlined in standard methodology (Sambrook *et al.,* 1989). After isolation of pure phage containing the gene of interest, mini- preparations or macro-preparation are made to isolate plasmid cDNA containing the gene of interest.

RT-PCR: To obtain the full-length sequence, 5' and 3' primers are designed based on the sequence information available in GenBank:

5' GCGCGGCCCAT GGAGTC 3' (SEQ.ID# 1)

is used as forward primer and

5' CTTTTCTGTGTTCCCAAGC AGAAC GTCAAT 3' (SEQ.ID#2)

is used as reverse primer. Total RNA from PBMC isolated from LGL leukemia patients and normal healthy individuals is used as a template for reverse transcriptase for making cDNA using either oligo(dT) primer or random hexamer primers. The PCR reaction mixture is heated to 95 ^{0}C for 2 min and then cycled 40 times at 95 ^{0}C for 30 sec, 60^{0}C for 45 sec, and 72 ^{0}C for 1.5 min. Finally, the reaction mixture is heated at 72 ^{0}C for 7 min and stored at 4 ^{0}C. The reaction product is electrophoresed in 1% agarose gels. For direct PCR, all the conditions are the same as above except that genomic DNA, isolated from PBMC, is used as a DNA template. PCR products are analyzed in l % agarose gel and the bands are excised and cloned into a TOPO-TA cloning vector (Invitrogen) and sequenced. The insert is subcloned into *EcoR1* sites of mammalian expression vector pcDNA3.1 to produce pcDNA3*S1P$_5$*.

Cell culture and transfection. HEK293 cells are grown in Dulbecco's modified eagle's medium supplemented with 10% fetal bovine serum. The cells are transiently transfected with a plasmid encoding HA-tagged Erk2 (HA-Erk2) and either pcDNA 3 *S1P$_5$* or pcDNA 3.1. Transfection is achieved by incubating the cells in 60 mm plates with plasmid/Lipofectamine complexes (2.1 μg total DNA/12 μl Lipofectamine) in serum-free medium for 5 hours. The DNA complexes are removed from the medium and the cells are starved overnight in serum-free medium and then used for experimentation.

Erk2 Kinase Assay. The serum-starved transiently transfected HEK293 cells are treated for 5 min preferably with either 1μM sphingosine-1-phosphate (S1P) or with 1μM lysophosphatidic acid (LPA). The cells are lysed in buffer containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 1% Triton X-100, 25 mM NaF, 5 mM sodium pyrophosphate, 20 mM ρ-nitrophenyl phosphate, 2 μg/mL leupeptin, and 100 μg/mL phenylmethylsulfonyl fluoride. HA-Erk2 is immunoprecipitated with the monoclonal antibody HA.11 (Convance, Richmond, CA). Half of the immunoprecipitate is used to determine Erk 2 activity and the other half is used for measuring Erk2 protein expression. For the kinase assay, immune complexes are incubated for 10 min at 30 ^{0}C in 40 μl of buffer containing 20 mM Hepes, pH 7.5, 10 mM $MgCl_2$, 1 mM dithiothreitol, 10 mM *p*-nitrophenyl phosphate, 40 μM ATP and 0.375 mg / mL myelin basic protein and 10 μCi of [γ-^{32}P] ATP (3000 Ci / mmol). The reaction is terminated with SDS-containing gel-loading buffer and the reaction mixtures are analyzed on 11% SDS-polyacrylamide gels. The gels are processed by autoradiography. The bands on the gels are quantitated with a PhosphorImager. Erk2 protein

in the immunoprecipitate is determined by immunoblotting with a polyclonal antibody to Erk2.

Examples

Referring now to FIG. 1, approximately 50 genes are up-regulated in LGL leukemia, with balanced differential expression of between about 7.8 and about 2.0. In addition, one EST is particularly noteworthy that is up-regulated in LGL leukemia with balanced differential expression of 3.0 (GenBank Accession number N47089). A clone containing this EST is sequenced. The total length of the EST is approximately 300 base pairs. A search using Blast shows 100% homology with another EST (GenBank Accession No. AF088014) named as homo sapiens full length insert cDNA clone YY85D04. No other information regarding this EST is found in the literature. No open reading frame is found within this sequence. Northern blot analysis confirms that a gene related to EST (GenBank ID No. N47098) is upregulated in majority of LGL leukemia patients.

Using the microarray screening method, one LGL leukemia patient is compared with one normal healthy individual. To show the same pattern in a larger sample of patents, Northern blot analysis is performed. Total RNAs, isolated from the PBMC of normal healthy individuals and LGL leukemia patients, are used in Northern blots. Initially, a 300 base pair cDNA fragment is used as a probe in initial experiments. Up-regulation of EST is observed in all the LGL leukemia patients when compared to the normal healthy individuals. This confirms the microarray results regarding EST expression. The probe hybridizes to a 2.2 kb transcript in the Northern Blots. (Fig-2).

An LGL leukemia library is constructed from the mRNA isolated from the pooled PBMCs of the seven LGL leukemia patients. This library is screened to obtain full-length sequence of the gene. Approximately 15 positive clones are selected and the larger clones are sequenced. The largest clone is 1500 bp in length. Analysis using Blast indicates that this gene has 85% homology with Rat *edg-8* (Im *et al,* 2000). All of the clones are missing 5'end of the gene. Blast search with htgs show 99% homology with the sequence present in chromosome 19. Based on the sequence information, primers are designed from the 5' end and from 3'end of the open reading frame of the gene. Three different products (1.5, 1.6, and 1.2 bp in length) are obtained using RT-PCR. These products are subjected to gel

electrophoresis and bands are excised, cloned into TOPO-TA cloning vectors and sequenced. The largest PCR product contains the entire open reading frame (FIG.3). The deduced amino acid sequence shows 85% homology with complete sequence of rat sphingosine 1-phosphate receptor *edg-8 and nrg-1.* (FIG. 4). Shorter bands are also identified. The shorter bands are excised, cloned, and sequenced. These clones are splice variants of sphingosine 1- phosphate receptor with deletions. They are herein termed "sphingosine 1-phosphate receptor -1" and "sphingosine 1-phosphate receptor-2" (FIG.5 & 6).

Expression of sphingosine 1-phosphate receptor is examined in different normal tissues by Northern blot analysis. It is found that *sppr* is expressed in several tissues such as brain, spleen and PBMCs. (FIG.7). Only trace amounts are detected in Jukat and CEM cell lines (data not shown).

To obtain a full-length sequence of the gene, an LGL leukemic cDNA library is constructed and screened using the EST probe. Approximately15 positive clones are selected and larger clones are sequenced. The BLAST search of the largest clone (1500 bp) indicates that this gene has strong homology with Rat *edg-8/Nrg-1*. However, all of the clones are missing the 5'end of the gene when compared to the rat gene. A BLAST search with the human genome shows 99% homology with a sequence present on chromosome19. Based on this sequence information, primers are designed from the 5' and 3'ends of the open reading frame of the gene.

Three different RT-PCR products (1.5, 1.6, and 1.2 bp) are obtained. These products are subjected to gel electrophoresis. The resulting bands are excised and cloned into TOPO-TA cloning vectors and then sequenced. The largest PCR product contains a complete open reading frame. The nucleotide sequence and the deduced amino acids are shown in Figure 8. The gene is designated as *S1P$_5$* (see below). The nucleotide sequence shows very strong homology with G-protein coupled receptors, especially with the endothelial differentiation genes (EDGs). When the deduced amino acid sequence of the full-length sequence is aligned with other members of the family using the CLUSTALW (multi sequence alignment) program, it is approximately 26 to 44% identical and 58 to 72% similar with EDGs at amino acid level (Table1). In addition, it shows 86% identity and 96% similarity with rat *edg-8* or rat *nrg-1* at amino acid level. (Figure 9, Table I). Transient transfection of HEK293 cells with this gene results in activation of Erk2 activity in response to sphingosine-1-phosphate but not

LPA, confirming that it is a sphingosine-1-phosphate receptor (Figure 10). Therefore, this gene is named $S1P_5$.

Samples from 30 LGL leukemia patients are tested for the presence of *S1P5* transcript by Northern blot analysis using full-length gene as a probe. Constitutive expression of *S1P5* transcripts is found in 24 samples (Fig. 11). In comparison *S1P5* transcripts are expressed at only trace levels in normal PBMC (N=12). After activation of normal PBMC the expression of *S1P5* is reduced to undetectable levels (Fig.12). Additionally, expression of two smaller bands is detected in samples from leukemic LGL by RT-PCR. Human $S1P_5$ transcripts are expressed mainly in normal brain, spleen, and PBMC and in trace amounts in lung, kidney and liver (Fig.13). Whereas expression of Edg-8 is observed only in brain and spleen of rat when Northern Blots are probed. Several cell lines are examined for the presence of $S1P_5$ transcript. Trace amounts of $S1P_5$ transcripts are identified in CEM and Jurkat cells (data not shown). All other cell lines tested are negative for *S1P5* transcript including MT2 (HTLV-I infected cell line) and MO-T (HTLV-II infected cell line), Moe7 (megakaryoblastic leukemic cell line) and U293 (human embryonic kidney cells).

Table 1. Identity and similarity between S1P5 and other members of the Edgs.

The deduced amino acid sequence of *S1P5* is aligned with the amino acid sequences of various members of Edgs. using the CLUSTALW program. Except for Edg 8 and nrg-1, all other sequences are from human. All the sequence information is obtained from GenBank.

Name of the gene	% Identity	% Similarity
hSiP5	100	100
rEdg-8*	87	96
rNrg-1	86	98
h Edg-1*	44	72
hEdg-5*	41	66
h Edg-3*	40	70
h Edg-6*	39	67
h Edg-2*	35	67
h Edg-4*	30	60
hEdg-7*	26	58

+ = Sphingosine 1- phosphate receptors

☆ = Lysophosphatidic acid receptors

Discussion

Leukemic LGL are resistant to Fas-induced apoptosis, in spite of over-expression of Fas and Fas-ligand (FasL) implying that the accumulation of circulating LGL can be due to dysregulation of apoptosis. The accumulation of circulating LGL in leukemic patients can also be due to clonal proliferation of LGL. In order to understand the molecular mechanisms involved in pathogenesis of LGL leukemia, microarray techniques are used to identify differentially expressed genes. Approximately 50 genes are identified that are up-regulated and 10 genes that are down regulated. Several ESTs are also identified which show differential expression. As a systematic study, one of the ESTs that is up-regulated in LGL Leukemia is characterized. The full-length gene is obtained by screening the LGL leukemia library and performing RT-PCR, which is 85% identical to the rat Sphingosine —1 Phosphate receptor. This gene belongs to G-protein coupled receptor super family and can act as a sphingosine-1-phosphate receptor. Several splice variants in LGL leukemia patients are also identified, and are named Sphingosine 1-phosphate receptor 1 and Sphingosine 1-Phosphate receptor 2. The deduced amino acid sequence of Sphingosine 1-Phosphate receptor with *rat edg-8 or nrg* shows 85% homology. It has seven transmembrane domains, which is a characteristic of GTP-coupled receptors. Thus, the Sphingosine-1 Phosphate is involved in the signal transduction from the sphingosine 1 –Phosphate in human.

Although the gene has lot of homology with other members of edg family, it is preferably named sphingosine-1-phosphate receptor ($S1P_5$) because it is mainly present in lymphocytes, brain and spleen, but not in endothelial cells.

Lysophosphatidic acid (LPA) and sphingosine 1-phosphate (S1P) mediate T cell function. Both LPA and S1P signaling pathways are implicated in cell proliferation, suppression of apoptosis, enhancement of cellular survival and T-lymphoma cell invasion. Although it has been suggested that S1P can act as an intracellular mediator, it has also been suggested that S1P acts as an extracellular ligand for cell surface receptors. Indeed several such receptors have been identified in a wide variety of tissues. For example, receptors Edg-1, -3, -5, -6 and –8, are specific for S1P, whereas Edg-2, -4, and –7 are LPA specific. In normal lymphocytes, there is differential constitutive expression of receptors for LPA and S1P. $CD4^+$ cells express predominantly Edg-4, while $CD8^+$ cells appeared to lack receptors for LPA and S1P as only traces of Edg-2 and Edg-5 are detected. Human T cell tumors express many Edgs for both LPA and S1P.

Rat edg-8 / nrg-1 is shown to be a sphingosine-1-phosphate receptor based on specific binding of radio-labeled S1P to cell membranes, inhibition of forskolin-induced cAMP accumulation, increased GTP binding ability and calcium mobilization studies. Even though these properties are adequate to classify *edg-8 / nrg-1* as a sphingosine-1-phosphate receptor, it seems surprising that this gene is different from other members of the human sphingosine-1- phosphate receptor family. For example, activation of EDG-1, -3, - 5 and –6 by S1P leads to activation of Erk1 / 2 and induction of cell proliferation. In contrast S1P inhibited serum-induces activation of Erk1 / 2 and also inhibits the cell proliferation in CHO cells expressing EDG-8. The reasons for these differences are not known and might be due to species variation. As shown herein, S1P activates Erk2 in transiently tranfected HEK293 cells while lysophosphatidic acid does not, suggesting that *S1P5* is a sphingosine-1-phosphate receptor and participates in sphingosine 1-phosphate mediated signal transduction. A computational model of the Edg-1 receptor predicts that Glu ^{121} is essential for interaction with S1P [21]. The S1P receptors Edg-1, -3, -5 and –8 as well as *S1P5* share such an anionic residue.

Leukemic LGL are antigen driven CTL that survive *in vivo,* at least in part, because of defective apoptosis. For example, leukemic LGL express both Fas and Fas-ligand, but are resistant to Fas mediated death. It is noteworthy that *S1P5* gene transcripts are down regulated after activation of normal T cells. Leukemic cells are activated T cells. Based upon the results disclosed herein, constitutive expression of *S1P5* transcripts represents dysregulated expression. This dysregulated expression of *S1P5* may participate in protection of leukemic LGL from apoptosis.

Note: The full-length sequence was deposited in GenBank (Accession No.AF331840) on December 22, 2000.

Throughout this application, various publications, including United States patents, have been referred to. The disclosures of these publications and patents in their entireties are hereby incorporated by reference into this application to more fully describe the state of the art to which this invention pertains.

While the invention has been described in terms of various preferred embodiments, those skilled in the art will recognize that various modifications, substitutions, omissions, and changes may be made without departing from the spirit of the present invention. Accordingly,

it is intended that the scope of the present invention be limited solely by the scope of the following claims.